$\mathbf{x}$	$\log_2(P(\mathbf{x}))$
...1...................1.....1....1.1.......1........1...........1.....................1.......11...	-50.1
......................1.....1.....1.......1....1.........1.....................................1....	-37.3
........1....1..1...1....11..1.1.........11.........................1..1.1..1...1.................1.	-65.9
1.1...1................1.......................11.1..1............................1.....1..1.11.....	-56.4
...11...........1...1.....1.1......1..........1....1...1.....1............1.........................	-53.2
..............1......1.........1.1.......1..........1............1...1......................1.......	-43.7
.....1........1.......1...1............1............1...........1......1..11........................	-46.8
.....1..1..1...............111...................1...............1.........1.1...1...1.............1	-56.4
.........1..........1.....1......1..........1....1..1...	-37.3
......1........................1..............1.....1..1.1.1..1...................................1.	-43.7
1.......................1..........1...1...................1....1....1........1..11..1.1...1........	-56.4
...........11.1.........1................1......1.....................1.............................	-37.3
.1..........1...1.1.............1.......11...........1.1...1..............1.............11..........	-56.4
......1...1..1.....1..11.1.1.1...1.....................1............1.............1..1..............	-59.5
............11.1......1....1..1............................1.......1..............1.......1.........	-46.8
....	-15.2
1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111	-332.1

Figure 4.10. The top 15 strings are samples from X^{100}, where $p_1 = 0.1$ and $p_0 = 0.9$. The bottom two are the most and least probable strings in this ensemble. The final column shows the log-probabilities of the random strings, which may be compared with the entropy $H(X^{100}) = 46.9$ bits.

Book by David MacKay

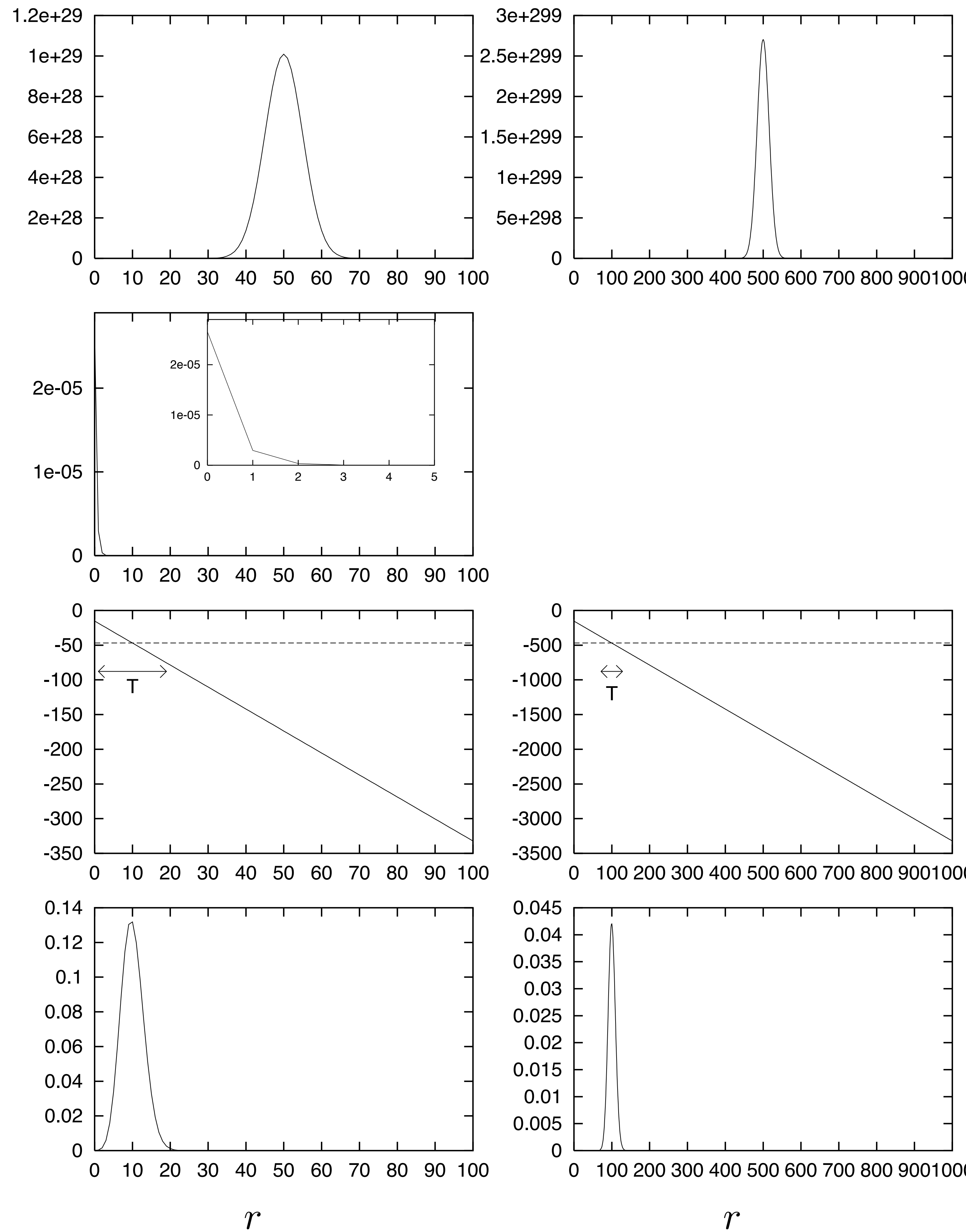
$N = 100$ | $N = 1000$

$n(r) = \binom{N}{r}$

$P(\mathbf{x}) = p_1^r (1 - p_1)^{N-r}$

$\log_2 P(\mathbf{x})$

$n(r)P(\mathbf{x}) = \binom{N}{r} p_1^r (1 - p_1)^{N-r}$



Book by David MacKay

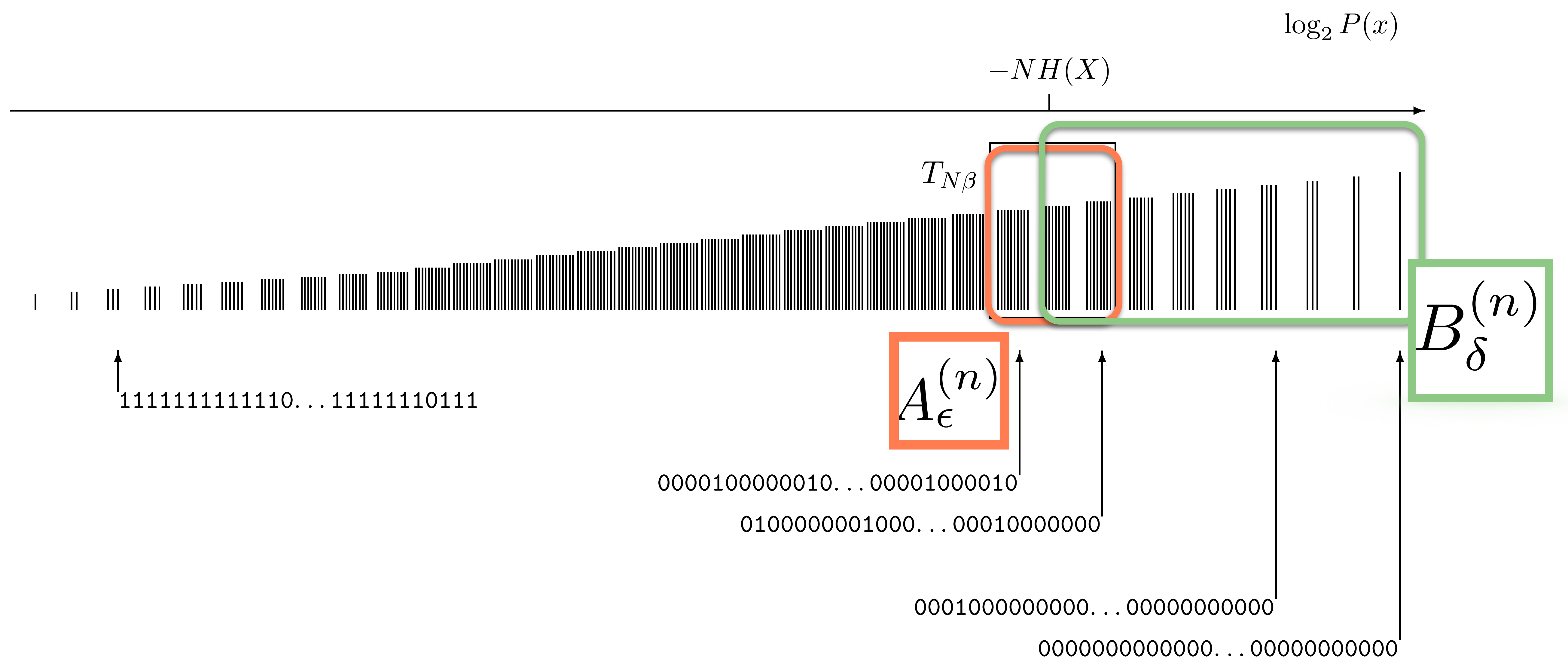


Figure 4.12. Schematic diagram showing all strings in the ensemble X^N ranked by their probability, and the typical set $T_{N\beta}$.

The 'asymptotic equipartition' principle is equivalent to:

Shannon's source coding theorem (verbal statement). N i.i.d. random variables each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss, as $N \to \infty$; conversely if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost.

at least $H - \epsilon$ bits. These two extremes tell us that regardless of our specific allowance for error, the number of bits per symbol needed to specify $\mathbf{x}$ is H bits; no more and no less.

Book by David MacKay